

IBWS: IST Bioinformatics Web Services

Achille Zappa¹, Mariangela Miele² and Paolo Romano^{2,*}

¹Department of Informatics, Systems and Telematics, University of Genoa, Via Opera Pia 13, I-16145, Genoa and ²Bioinformatics, National Cancer Research Institute of Genoa, Largo Rosanna Benzi 10, I-16132, Genoa, Italy

Received February 12, 2010; Revised April 23, 2010; Accepted May 4, 2010

ABSTRACT

The Bioinformatics group at the National Cancer Research Institute (IST) of Genoa has been involved since many years in the development and maintenance of biomedical information systems. Among them, the Common Access to Biological Resources and Information network services offer access to more than 130 000 biological resources, like strains of micro-organisms and human and animal cell lines, included in 29 collections from some of the most known European Biological Resource Centers. An Sequence Retrieval System (SRS) implementation of the TP53 Mutation Database of the International Agency for Research on Cancer (Lyon) was made available in order to improve interoperability of this data with other molecular biology databases. 'SRS by WS (SWS)', a system for retrieving information on public SRS sites and for directly querying them, was also implemented. In order to make this information available through application programming interfaces, we implemented a suite of free web services (WS), called the 'IST Bioinformatics Web Services (IBWS)'. A support web site, including a description of the system, a list of available WS together with help pages, links to corresponding WSDLs and forms for testing services, is available at <http://bioinformatics.istge.it/ibws/>. WSDL definitions can also be retrieved directly at <http://bioinformatics.istge.it:8080/axis/services>.

INTRODUCTION

Nowadays, biological data is spread through heterogeneous information systems that are distributed over the Internet. Among current information and communication technologies, workflow management systems, in connection with web services (WS), seem to be the most

promising ones (1). For them to have the greatest possible impact on biological data integration and analysis, it is important that all major databases and information systems, in the many biology domains, be made available through standard programmatic interfaces.

The Bioinformatics group at the National Cancer Research Institute (IST) of Genoa has been involved since many years in the development and maintenance of biomedical information systems, some of which are unique in their domain.

The Common Access to Biological Resources and Information (CABRI) project was funded by the European Union from 1996 to 1999 in the sphere of the V Framework Programme. It led to the implementation of a 'one-stop-shop' (<http://www.cabri.org/>) for biological resources of high quality maintained by European Biological Resource Centers that agreed to adopt common high quality standards for the management of materials and of data (2).

The CABRI site currently is a well known information source on biological resources, with more than 1 750 000 hits/month in December 2009 and January 2010 (a hit corresponding to any file requested to the server, excluded images and requests submitted by known robots or by users in the local network). These statistics was computed by using the AWStats log file analyzer (<http://awstats.sourceforge.net/>) that is able to detect and to remove visits from more than 300 distinct robots.

The TP53 Mutation Database of the International Agency for the Research on Cancer (IARC) is the biggest and most detailed database of mutations described in literature on the TP53 human gene and related protein (3). In the IARC web site (<http://www-p53.iarc.fr/>), queries can only be executed on-line and imply a human interaction. Moreover, some data sets are not searchable on-line and mixed queries, involving more data sets, are not possible. In order to improve accessibility and interoperability with other databases, we implemented the IARC database in an SRS site (4).

A list of public SRS sites is maintained by BioWisdom Ltd at <http://downloads.biowisdomsrs.com/publicsrs.html>. Each of these sites can be queried through its web

*To whom correspondence should be addressed. Tel: +39 010 5737288; Fax: +39 010 5737295; Email: paolo.romano@istge.it

interface, but their contents are not available through a programmatic interface. Moreover, users are usually obliged to take note of existing sites and relative database implementations, in order to drive their activity to a special site. Network issues and server unavailability may make the use of the sites difficult. SRS by WS (SWS) is a system that allows to query biological databases available in the list of public sites and to return results in a simple text-only format (5). It allows essential information to be retrieved both for the sites, such as their current availability and lists of installed databases and tools, and for the databases, including where they have been implemented, their relative sizes and versions.

In order to make CABRI catalogues, the IARC TP53 Mutation Database and the SWS system available through Application Programming Interfaces (APIs), for the good of programmers, we implemented a suite of free WS, called the 'IST Bioinformatics Web Services (IBWS)', that are described in the 'Results' section of this article.

MATERIALS AND METHODS

Data sources

CABRI catalogues describe more than 130 000 biological resources, namely strains of micro-organisms, plasmids, phages, human and animal cell lines, and plant cells and viruses, which are included in 29 collections maintained by European Biological Resource Centers. CABRI catalogues also include cross-reference links to other databases, like the EMBL data library (that contains back links to CABRI) and Medline. CABRI catalogues underwent a careful analysis and comparison before their inclusion in the CABRI site. This led to the definition of common data sets and formats.

The IARC TP53 Mutation Database includes somatic mutations (mutations and related bibliographic references, mutation prevalence and prognostic value, gene variations and polymorphisms), germline mutations (both data and bibliographic references), mutant functions and cell line status. Release 14, issued in November 2009, includes 26 597 somatic mutations whose description has been derived from 2198 papers which are included in Medline. Reference vocabularies and standardized annotations are used extensively for the description of the mutation, tumour site, type and origin, and for literature references. Examples of the former are International Classification of Diseases for Oncology (ICD-O) and SNOMED nomenclatures.

In our SRS implementation, we aimed at exploiting all features of the database. Terms included in the controlled vocabularies have been properly indexed and can also be used from within the SRS extended query form, thus allowing for a data-driven search. SRS fields have been formatted according to the type of the corresponding database field at IARC. They can therefore be searched by using checkboxes (when an enumeration is used), numeric ranges (for numerical values), text fields and multi select boxes having the items of controlled

vocabularies as reference lists from which to make the selection.

SRS internal links allow to retrieve data from one data set by imposing conditions on another. For example, retrieve all papers describing a given mutation type or all mutations described by papers appeared in a given journal. HTML links from the bibliographic references data set to PubMed at the NCBI allow direct access to corresponding information, either abstracts or full texts.

As said, SRS implementations of CABRI catalogues and of the IARC TP53 Mutation Database are respectively available at <http://www.cabri.org/> and at <http://srs.o2i.it/>. All IBWS related to CABRI and TP53 make direct access to these SRS implementations by either using local or remote invocations, the latter involving the use of the `wgetz` script and the former using `getz` instead.

CABRI catalogues are not updated regularly: some of them are updated many times per year, while some others have not been updated recently. They are assigned a version number that includes the year and a progressive number. For example, version 2009.5 is the fifth update in 2009. IBWS gives always access to the most recent release.

The IARC TP53 Mutation Database is updated annually, usually in November. Our SRS implementation is updated with some delay because of changes that may occur in the data structure. Release 14 was issued in November 2009 and its SRS version was made available in January 2010. Information is included in 12 interlinked distinct data sets.

The list of SRS sites that is made available by BioWisdom Ltd is updated daily. With the same frequency, SWS checks the list, extracts data and stores it into a local relational database, the `srddb`, that includes tables for databases, sites and implementations. According to requests made by users, IBWS give access to `srddb` and/or to SRS sites. Currently there is no web interface for querying `srddb`.

SWS can be invoked by specifying the name of the databank to be queried and query terms. It then automatically chooses the best site, performs the query and returns the complete results. Users can also specify the following information: the SRS site to be queried, the fields where the information must be searched and the desired output fields.

When querying, terms that must be searched for and the database to search are mandatory input. The site, instead, can be omitted. In that case, SWS identifies the best one by selecting, among those that are active, the site where that specific library has the greatest number of entries and, when more sites have the same number, the most recent version of SRS (this function is limited to SRS versions 6 and 7). Further optional parameters allow to define which fields must be queried and which one must be returned.

Deployment of IBWS

All WS in IBWS have been deployed by using SoapLab, a SOAP-based analysis web service providing a programmatic access to local, command-line applications, like the EMBOSS software, and to the contents of ordinary web pages (6,7). The only requirements of SoapLab are

the Apache Tomcat servlet engine with the Axis SOAP toolkit, a Java Virtual Machine and, optionally, perl and MySQL.

Once the server has been installed, new WS are deployed (i.e. made available through the programmatic interface) by defining interface parameters that specify the task to be performed, either a local execution command or a remote URL invocation. Each IBWS makes then reference to one of various scripts (mainly written in perl or Unix shell, in our case) that actually execute the requested operation, e.g. run getz, or launch a call to wgetz or make access to srsdb. Such definitions are written in the AJAX Command Definition language and must be converted to XML before they can be used by SoapLab.

RESULTS

Access to IST Bioinformatics Web Services

A support web site, including a description of the system, and a list of available WS, together with help pages, links to corresponding WSDLs, examples of their usage, and a simple test form, is available at <http://bioinformatics.istge.it/ibws/>. The web page showing usage of the three TP53 general access WS is provided as Supplementary Data, file [tp53_service_web_form.pdf](#). WSDLs can also be retrieved directly by software tools from <http://bioinformatics.istge.it:8080/axis/services>.

A schema of the system, including links between its components and ways to interact with it by APIs and by web pages is shown in Figure 1.

IBWS can be accessed through any WSDL-SOAP compliant software, including the well-known Taverna Workbench (<http://www.taverna.org.uk/>) (8). Since IBWS were implemented by using SoapLab, any SoapLab-enabled client should work for them too.

In order to support the use of IBWS, we also developed some simple template client software for accessing them by using PHP and perl (Supplementary Data, files [ibws_php_client.pdf](#) and [ibws_perl_client.pdf](#)). The client software written in PHP uses NuSOAP – SOAP Toolkit for PHP (<http://sourceforge.net/projects/nussoap/>). The client software written in perl uses SOAP-Lite (SOAP Toolkit for Perl, <http://search.cpan.org/~byrne/ SOAP-Lite-0.60a/>). All templates can be downloaded from the IBWS support web site.

We plan to accumulate more sample scripts: for special needs, readers are welcome to email to authors. Further user support is available and readers are encouraged to contact the group with problems, comments and suggestions (the support site for contact information). They can also get in touch with the group leader through some of the many social networks available, including, e.g. myExperiment (<http://www.myexperiment.org/>), where some workflows making use of IBWS are available (9), and BioCatalogue (<http://www.biocatalogue.org/>), where all IBWS are registered (10).

IBWS currently include three main groups, respectively referring to CABRI catalogues, to data sets of the IARC TP53 Mutation Database and to SWS. Overall, 66 WS are available through the IBWS portal.

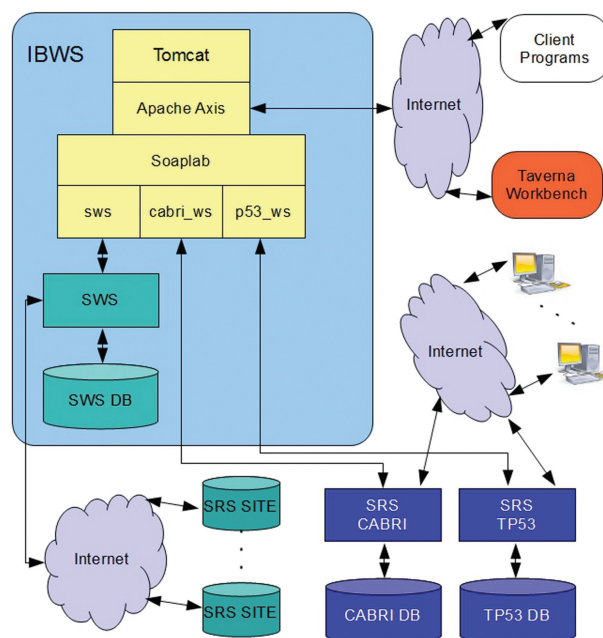


Figure 1. A schema of the IBWS system, including links between its components and ways to interact with it by APIs and by Web pages. Users may interact with the APIs through Internet by using their preferred client programs. Interaction is managed by an Axis/Tomcat implementation that requests SoapLab to launch the appropriate task. SWS-related services may either make access to the SWS database (srsdb) or query public SRS sites remotely. Both CABRI- and TP53-related services make access locally to their respective SRS implementation. These may also be directly queried by users through any browser.

IBWS devoted to CABRI

CABRI WS have been designed to offer the same functionalities of the CABRI Simple Search interface (<http://www.cabri.org/CABRI/srs-doc/>). They therefore allow to query catalogues by the name of the biological resource (this can be the scientific name when searching for micro-organisms' strains), by its identifier (usually a collection number) or free text, in which case all textual information is searched. Searches usually return the identifiers of the biological resources and these on their turn can be used to retrieve full records. Two types of services were therefore implemented: one that allows to search for a feature of interest (e.g. name, origin, host) and returns the identifiers of matching records, and the other that allows to search by identifier and returns full records.

Distinct services have been developed for each of the main types of biological resources in the CABRI system and for all types together. This can be achieved by querying the Interconnected Biological Resources Database (IBRD) where names and identifiers for all biological resources are gathered. Further WS are available for a standardized access to any CABRI catalogue.

See Table 1 for the names of WS. For further help on their use, see the support web site.

IBWS devoted to TP53

IBWS devoted to TP53 allow for the retrieval of either database identifiers or complete records from any of the

Table 1. IBWS related to CABRI catalogues, grouped by application and domain

Application	Domain	WS
cabri	Bacteria	getBacteriaIdsByName, getBacteriaIdsByProperty, getBacteriaById
	Filamentous fungi	getFungiIdsByName, getFungiIdsByProperty, getFungiById
	Yeasts	getYeastIdsByName, getYeastIdsByProperty, getYeastsById
	Human and animal cell lines	getCellLineIdsByName, getCellLineIdsByProperty, getCellLinesById, getCellLineByName, getCellLineNames
	Plasmids	getPlasmidIdsByName, getPlasmidIdsByProperty, getPlasmidsById
	Phages	getPhageIdsByName, getPhageIdsByProperty, getPhagesById
	IBRD	getResourceIdsByName, getResourcesById
cabri_services	General access	getCabriEntriesByField, getCabriFieldsByField, getCabriIdsByField

Table 2. IBWS devoted to the access to the TP53 Mutation Database, grouped by application and domain

Application	Domain	WS
o2i	Somatic mutations data set	getP53CodonNumberById, getP53CodonNumbersWtaaMutaaById, getP53MutaaByIds, getP53MutationsByIds, getP53PubMedIdByIds, getP53PubMedIdByRefId, getP53RefIdByIds, getP53WtaaByIds, getP53SampleNames, getP53MutationIdsBySampleName
	Mutation functions data set	getP53MutationFunctionEntry, getP53MutationFunctionEntryById, getP53MutationFunctionIdsByCodonNumber, getP53MutationFunctionIdsByMutaa
	Cell line status data set	getP53CellLinesStatusEntryById
	All data sets (according to availability of required fields)	getCodonNumberByIds, getIdsByCodonNumber, getIdsByMutaa, getMutaaByIds, getP53MutationsBySampleName, getP53MutationIdsByType, getP53MutationIdsByTumorOrigin, getP53MutationIdsBySpliceSite, getP53MutationIdsByProperty, getP53MutationIdsByMetastasisLocalization, getP53MutationIdsByIntron, getP53MutationIdsByCodonNumber, getP53MutationIdsByCpgSite, getP53MutationIdsByEffect, getP53MutationIdsByExon
p53_getz_services	General access	getP53EntriesByField, getP53FieldsByField, getP53IdsByField

TP53 data sets available in the SRS implementation. Also in this case, two types of services were implemented, allowing to search the various data sets either for a specific feature and returning IDs, or for an ID and returning full records. Moreover, the tp53_mutation data set, which is related to somatic mutations, can be searched by many characteristics, like exon or intron number where the mutation occurs, effect of the mutation on the coding DNA, mutation type, tumour origin, and occurrence of the mutation in a splice site or in a CpG island.

Distinct services have been developed for each data set in the TP53 Mutation Database. Further WS are available for a standardized access to any data set, in which case the name of the data set must of course be specified.

See Table 2 for the name of all WS. For more help, refer to the support web site.

IBWS devoted to SWS

IBWS devoted to SWS allow to query biological databases included in public SRS sites and to return results in a simple text-only format. It allows essential information to be retrieved both for the sites, such as their current availability and lists of installed databases and tools, and for the databases, including where they have been implemented, their relative sizes and versions. It also allows to query selected systems by specifying the name of the databank to be queried and query terms. In this case, it may automatically choose the best site, perform the query and return complete results. Users can also specify the site to be queried, the fields where the information must be searched and the desired output fields.

SWS currently includes five services. getDBs retrieves acronyms of all libraries (databases) that are available in

a specified site. `getSites` retrieves acronyms of all SRS sites that include a specified library. `getImplementations` retrieves all implementations of a specified library. These services do not actually query any SRS site. `querySWS`, allows to actually perform queries on a specified library. Finally, `testSites` tests if a defined site is actually active at the time of its invocation.

Making use of IST Bioinformatics Web Services

One simple way of making use of IBWS is to develop workflows by using the Taverna Workbench, a well known, open source workflow management system (Taverna SourceForge site at <http://taverna.sourceforge.net/>). Taverna supports the rapid development of complex workflows by combining diverse services or simpler workflows and provides a practical and useful way to document and share workflows and their design. Some example workflows that leverage on IBWS and were developed by using Taverna are available at the myExperiment site (<http://www.myexperiment.org/>) (9). They were designed to demonstrate how IBWS can be interconnected and linked with other WS. Examples include, workflows for making a simple query to SWS and for retrieving known mutations of the TP53 gene that are present in any of the human cell lines that are available from CABRI catalogues.

Interconnection among IBWS can be achieved mainly through the utilization of identifiers. Indeed, as said, IBWS comprise two types of services: those providing identifiers of records matching a search criteria (that normally refer to any contents of the databases) and those returning complete records matching a given

identifier. So, one effective way to interconnect services is to search for specific features, retrieve related records' identifiers and request complete entries. Identifiers resulting from many searches can of course be combined as needed.

The complete diagram of a first example workflow, showing the usage of IBWS, is provided as additional material as `workflow_diagram_1.pdf`. The workflow can also be downloaded from myExperiment at the following URL: <http://www.myexperiment.org/workflows/36>. In Figure 2, a simplified diagram of the same workflow, showing its main elaboration steps without entering too much into details, is presented.

The goal of this workflow is to retrieve all TP53 mutations that are known to be present in cell lines available from CABRI catalogues. This goal implies the utilization of various IBWS for querying both CABRI cell lines catalogues, the TP53 somatic mutations data set and eUtils (services offered by the NCBI, <http://eutils.ncbi.nlm.nih.gov/>). First, the TP53 somatic mutations data set is searched to retrieve all sample names (i.e. cell line names) that were found to present TP53 mutations (the `getP53SampleNames` service is used). Second, all cell line names from one or more of the CABRI cell lines catalogue are retrieved (the `getCellLineNames` service is used here). These names can possibly be used to retrieve complete entries of cell lines from the same library/ies using `getCellLineByName`. Results from the first two services are merged, filtered and adapted so that they can be used to retrieve mutation identifiers, again from the TP53 somatic mutations data set (the `getP53MutationIdsBySampleName` and `getP53MutationIdsBySampleName` services are used to this end). Then, PubMed IDs of papers describing those mutations are retrieved by using mutation identifiers (the `getP53PubMedIdByIds` service is used). Finally, these identifiers can be used to obtain relevant PubMed records. All resulting outputs are then concatenated to form the final output (an example output is presented as a Supplementary Data, file `workflow_output_sample.pdf`).

The diagram of a second example is also provided as additional material as `workflow_diagram_2.pdf`. This workflow can also be downloaded from myExperiment at the following URL: <http://www.myexperiment.org/workflows/1049>. In this case, the goal is to extract and plot information about the distribution of some characteristics of TP53 mutations given a morphology type. This goal can be achieved by querying first the TP53 somatic mutations data set with the aim of retrieving an ordered list of topographies and related mutation identifiers. These identifiers can then be used to retrieve complete entries (the `getP53MutationsByIds` service is used) and lists of related effects, codon of occurrence, mutation type and topology (the `getP53FieldByField` service is used for each of these features, separately). Resulting lists can then be used as input to the Rshell processor (11), a Taverna plugin that, given an appropriate R script, may launch, either locally or remotely, an implementation of Rserve (12) to create distribution plots of these characteristics.

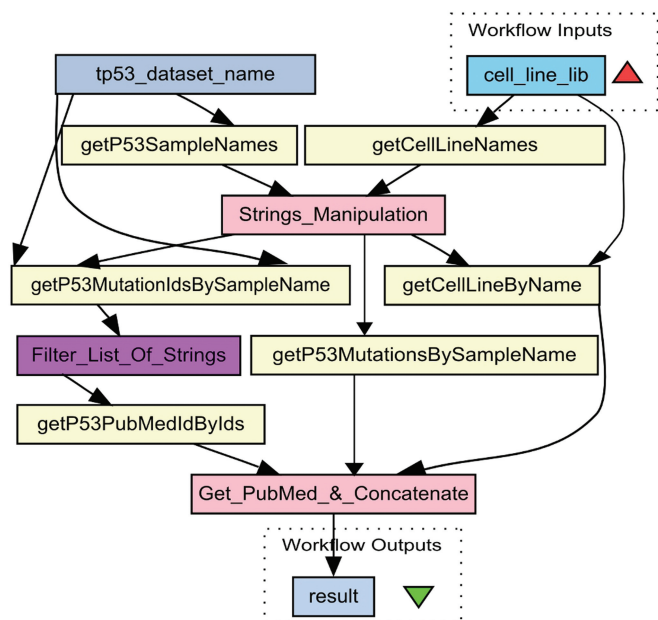


Figure 2. Simplified workflow demonstrating the use of IBWS for interconnecting TP53 Mutation Database, CABRI catalogues and PubMed. Only main processing steps are shown, while string manipulations and other accessory elaborations are grouped. See description in the text.

CONCLUSION

In this article, we have presented IBWS, a suite of free WS that allow to make access through a programmatic interface to some information systems that were developed at IST, the National Cancer Research Institute of Genoa, and are not available in any other web site. IBWS are documented in a purpose support web site, where the list of all WS, relative help pages, links to corresponding WSDLs and forms for testing services are presented together with an overall description of the system. IBWS have been developed by using standard tools and are therefore likely to be easily invoked by any software that is compliant with those standards, including the Taverna Workbench. Template client scripts have been made available for the PHP and perl scripting languages and can be used to check invocation methods and proper use of IBWS.

The main advantage offered by IBWS relates of course to the possibility of accessing through SOAP APIs a set of unique archives, that otherwise can only be queried manually. From this, well known consequences derive: possibility to query these data sources, to interconnect with other information systems, to build up workflows, even complex ones, and, finally, to implement automated data analysis processes.

There are of course some limitations and we are planning to implement new features aimed at solving some of them. The lack of detailed information and examples about the parameters, their appropriate values and the interrelationships between them may be an impediment to using many bioinformatics tools in an effective way. In order to address this issue, we plan to add extensive metadata to IBWS. This includes both adding operation and data structure documentation in the support site and providing new services to obtain information about meaning and use of parameters, as well as lists of their allowed values. This would especially be useful for some of IBWS, namely those related to the TP53 Mutation Database, which may leverage from the standardization of data that is inserted into the database according to controlled vocabularies.

In accordance with WS best practices, the RPC/encoded services will be migrated to the WS-I (<http://www.ws-i.org/>) recommended Document/literal style in a new SoapLab2 platform structure. Also, we are taking into account the possibility of implementing REST-style services that will be updated to provide service description documents in accordance to the Web Application Description Language (<https://wadl.dev.java.net/>) specification to allow automated utilization of these services.

Furthermore, we plan to develop new WS aimed at exposing information from two further information systems, namely the Cell Line Data Base and its associated Cell Line Integrated Molecular Authentication database (13).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online

ACKNOWLEDGEMENTS

Our system is partially based on open source. We wish to acknowledge all software developers, database administrators, data curators and users at their Institute and elsewhere, who have provided extremely valuable feedback and support throughout.

FUNDING

Italian Ministry of Education, University and Research (MIUR), projects 'Oncology over Internet (O2I)', 'Laboratory of Interdisciplinary Technologies in Bioinformatics (LITBIO)'; Italian Ministry of Health, project 'Italian Network for Oncology Bioinformatics (RNBIO)'. Funding for open access charge: Project 'Rete Nazionale Bioinformatica Oncologica', Italian Ministry of Health.

Conflict of interest statement. None declared.

REFERENCES

- Romano, P. (2008) Automation of in-silico data analysis processes through workflow management systems. *Brief. Bioinformatics*, **9**, 57–68.
- Romano, P., Kracht, M., Manniello, M.A., Stegehuis, G. and Fritze, D. (2005) The role of informatics in the coordinated management of biological resources collections. *Appl. Bioinformatics*, **4**, 175–186.
- Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S.V., Hainaut, P. and Olivier, M. (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.*, **28**, 622–629.
- Marra, D. and Romano, P. (2007) Integrating mutation data of the TP53 human gene in the bioinformatics network environment. In *Proceedings of the First International Conference on Bioinformatics Research and Development BIRD '07*, March 12–14, 2007, Berlin. Springer Lecture Notes in Bioinformatics LNBI 4414, Springer Berlin, Heidelberg. pp. 453–463.
- Romano, P. and Marra, D. (2008) SWS: accessing SRS sites contents through Web Services. *BMC Bioinformatics*, **9**(Suppl. 2), S15.
- Senger, M., Rice, P. and Oinn, T. (2003) Soaplab — a unified Sesame door to analysis tools. In Cox, S.J. (ed.), *Proceedings of the UK e-Science All Hands Meeting 2003*. The Engineering and Physical Sciences Research Council (EPSRC), Swindon, pp. 509–513.
- Senger, M., Rice, P., Bleasby, A. and Uludag, M. (2009) Soaplab: open source web services framework for Bioinformatics programs. *The 10th Annual Bioinformatics Open Source Conference. Satellite Workshop of the Joint 17th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2009) and 7th European Conference on Computational Biology, ECCB 2009*, p. 168, June 27 – 28, Stockholm, Sweden.
- Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M., Li, P. and Oinn, T. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Research*, **34**, W729–W732.
- De Roure, D., Goble, C. and Stevens, R. (2009) The design and realisation of the myExperiment virtual research environment for social sharing of workflows. *Future Gener. Comput. Syst.*, **25**, 561–567.
- Goble, C.A., Belhajjame, K., Tanoh, F., Bhagat, J., Wolstencroft, K., Stevens, R., Nzuobontane, E., McWilliam, H., Laurent, T. and Lopez, R. (2009) BioCatalogue: a curated web service registry for the Life Science community. *Nature Precedings*. <http://dx.doi.org/10.1038/npre.2009.3132.1>.

11. Wassink,I., Rauwerda,H., Neerincx,P.B.T., van der Vet,P.E., Breit,T.M., Leunissen,J.A.M. and Nijholt,A. (2009) Using R in Taverna: RShell v1.2. *BMC Res. Notes*, **2**, 138.
12. Urbanek,S. (2003) Rserve – a fast way to provide R functionality to applications. In Hornik,K., Leisch,F. and Zeileis,A. (eds), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20–22, 2003. Vienna, Austria.
13. Romano,P., Manniello,A., Aresu,O., Armento,M., Cesaro,M. and Parodi,B. (2009) Cell Line Data Base: structure and recent improvements towards molecular authentication of human cell lines. *Nucleic Acids Res.*, **37**, D925–D932.